ItronUtilityWeek | 2014
resourcefulness in action

PERFORMANCE METRICS AND OBJECTIVE TESTING METHODS FOR ENERGY BASELINE MODELING SOFTWARE

STREAMLINING M&V THROUGH AUTOMATION AND ANALYTICS

Jessica Granderson, PhD, Lawrence Berkeley National Lab

jgranderson@lbl.gov

PRESENTATION OUTLINE

» Motivation and Background
» Approach
» Key Results
» Looking Forward
» Q&A

MOTIVATION

» **High level goal**: Enable the industry to harness emerging tools and devices to conduct M&V at dramatically lower cost, with comparable or improved accuracy

» LBNL and QuEST are growing a body of research in streamlining, automation, accuracy and uncertainty in M&V

- Past and current support from CEC, PGE, and DOE-BTO

AUTOMATED M&V IS AN EMERGING CAPABILITY IN TODAY'S MORE ADVANCED ANALYTICAL TOOLS

Automated M&V is beginning to be offered in building information technologies, analytical software tools

Baselines are automatically created using historic interval meter data (system level or whole-building) and weather data feeds

Regression, NN, Bin models most common

User enters the date of EEM implementation, savings automatically calculated

WHAT QUESTIONS ARE BEING ASKED?

» How can I determine whether a given model or commercial tool is robust and accurate?

» What repeatable test procedures can be used to evaluate model and tool performance, and which metrics provide critical performance insights?

» How can I compare and contrast proprietary tools and ‘open’ modeling methods for M&V?

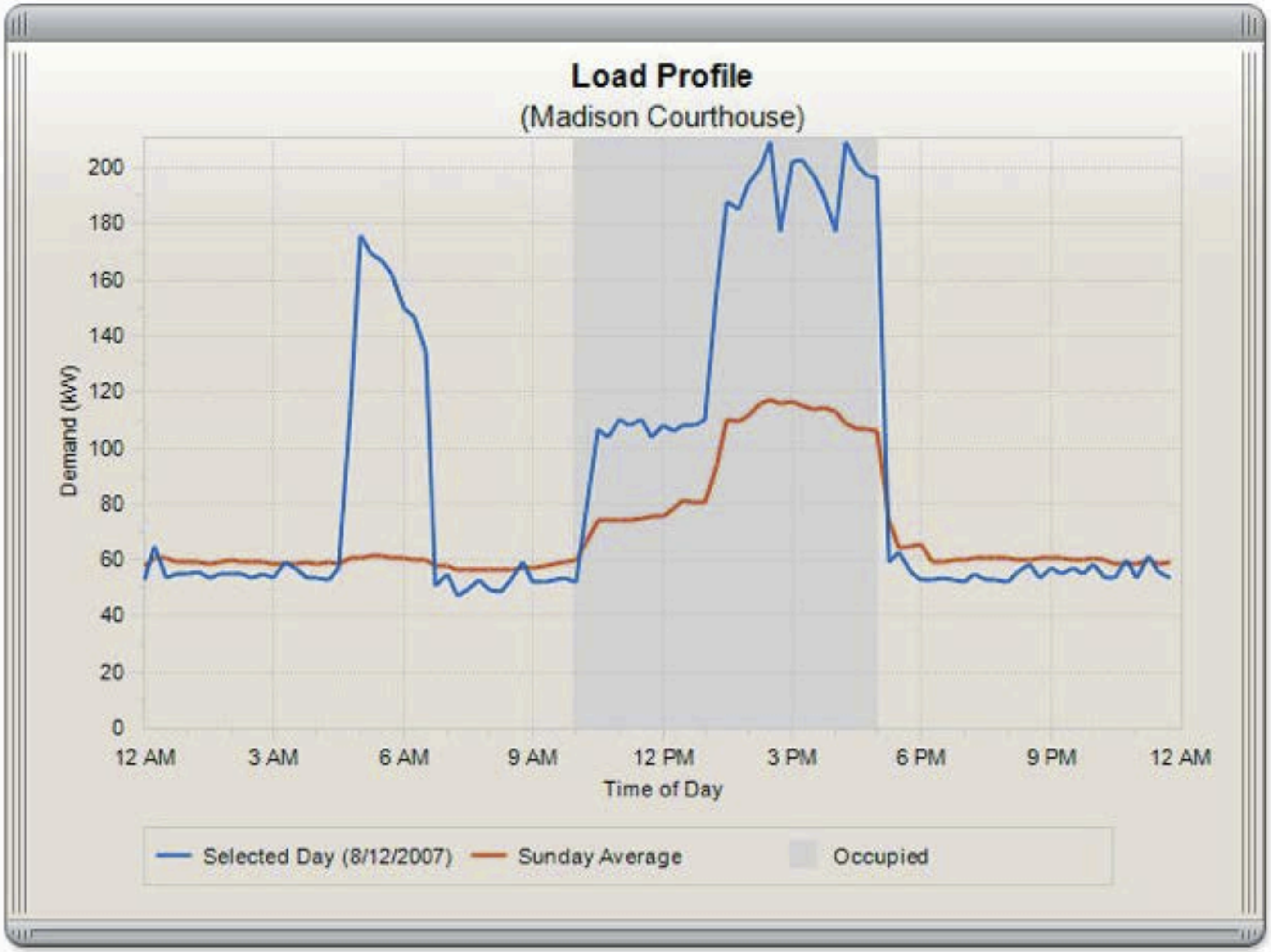
» How can we reduce the time and costs necessary to quantify gross savings?

» Can I use a whole-building approach for my programs and projects?

*In contrast to post-project, verification questions – how much was saved, what was the uncertainty?

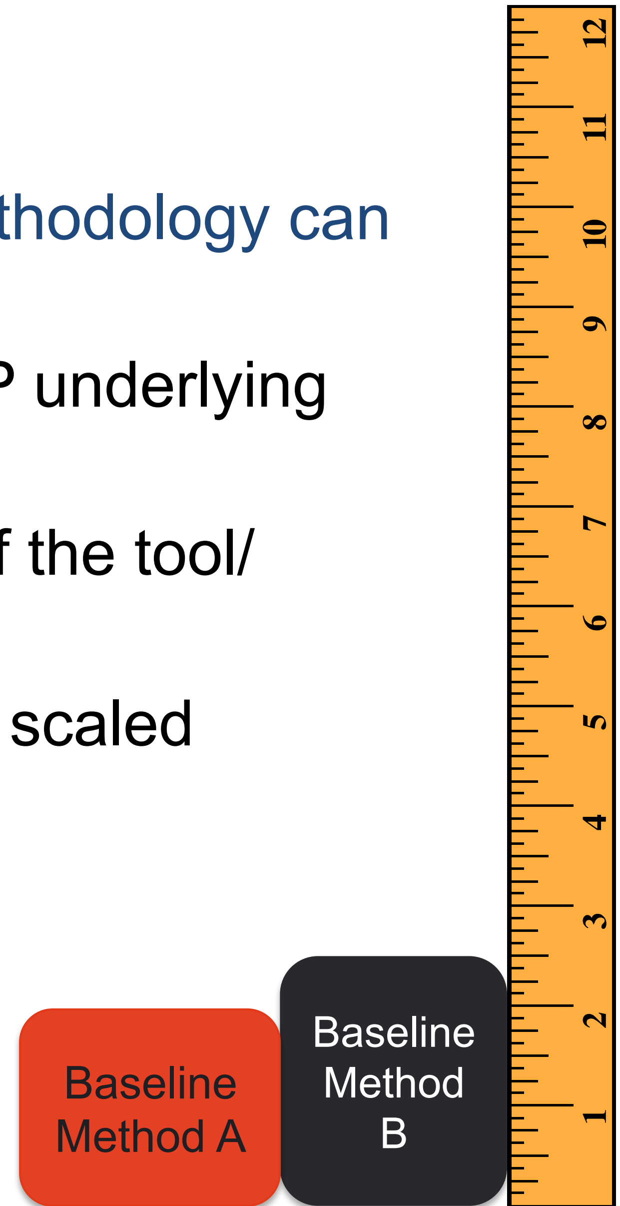
WHAT IS AN ENERGY BASELINE?

Example: energy anomaly detection of waste in real time



R&D TO ASSESS M&V/BASELINE PERFORMANCE ACCURACY

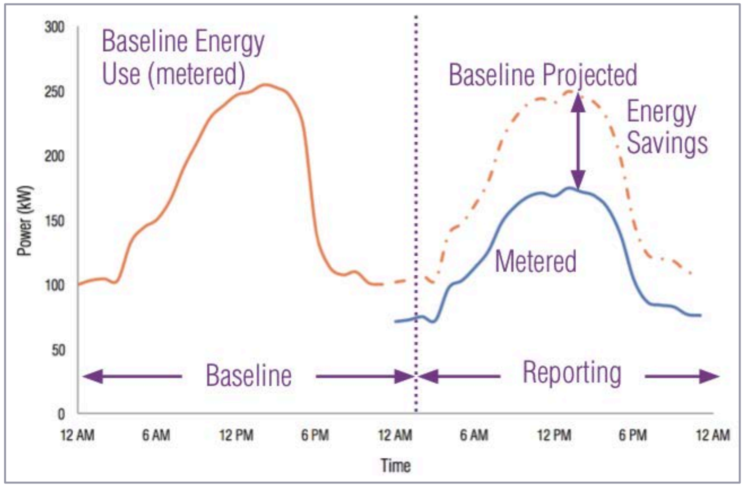
- » Objective performance assessment methodology can provide a win/win
  - Allow vendors to retain proprietary IP underlying the algorithms
  - Allow users to gauge performance of the tool/ approach
  - Give industry confidence needed for scaled deployment, widespread adoption



Approach: Objective Performance Testing Methodology

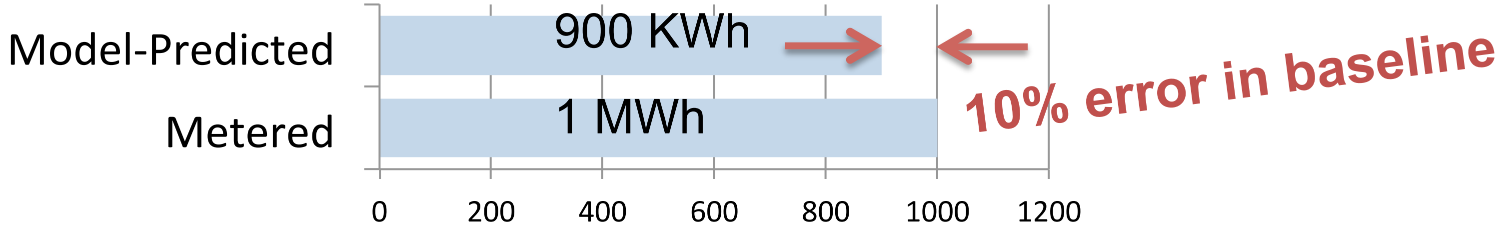
HOW ACCURATE IS THE BASELINE MODEL?

M&V Use Case



Error in reported savings is proportional to error baseline projection

Error = **% difference** between total **metered** energy use, total **model-predicted use**



HOW DO WE ASSESS THESE ERRORS?

Baseline Model

Test Data: Many buildings, metered data

Model

- **Split data set** into training & prediction period.
- **Train the model** by showing data, hiding prediction-period data;
- **Generate post-period predictions**

Compare

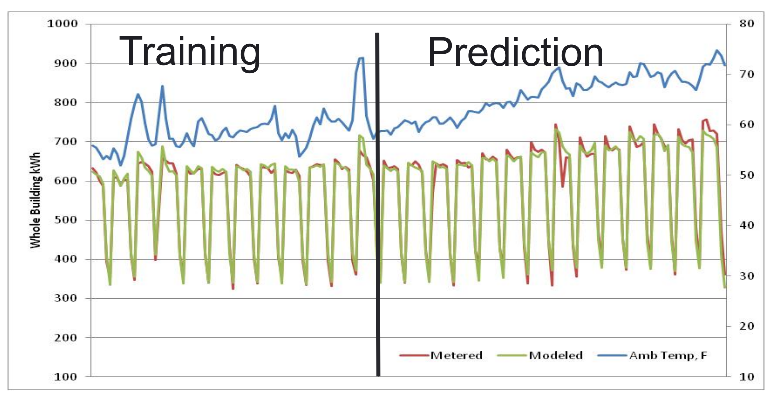
Compare predicted data to actual data that was 'hidden' from model to quantify error.

Repeat for many buildings

Assess

Calculate Performance Metrics, e.g. %Error, R^2, CV(RMSE) …

Outdoor Temp
Metered
Modeled



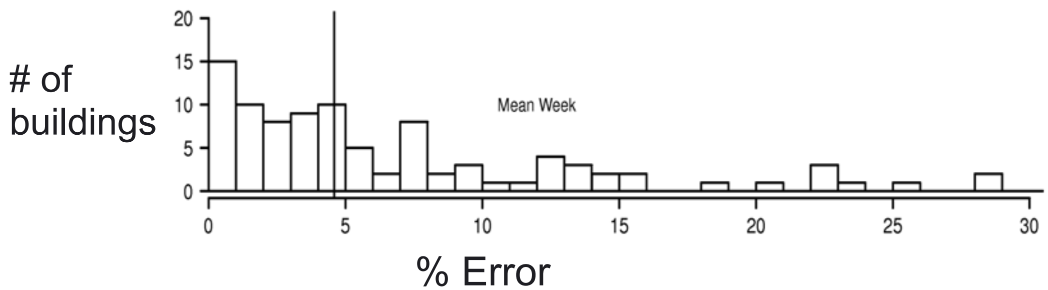
Key Results

MEDIAN ERROR OF 5% ACROSS 100'S OF BUILDINGS

» 5 models: change-point and more sophisticated regression models, interval and monthly data

» 12 months training (pre) and 12 months prediction (post)

» Median error was ~5%; Mean error was ~8%



» Consider trade-offs between reducing cost/full automation, and highest accuracy (engineer involved)

HOW DEEP DO SAVINGS HAVE TO BE?

Percentiles of Errors

Model	10%	25%	50%	75%	90%	Mean
Mean Week	0.82	2.21	4.82	9.63	19.42	8.40
Monthly CDD and HDD	0.69	2.09	4.53	10.03	19.38	8.46
Day, Time, and Temperature	0.69	2.17	4.51	9.26	19.41	8.42
Day and Change Point	0.73	2.02	4.70	9.22	18.84	8.24
Time of Week and Temperature	0.82	2.21	4.82	9.63	19.42	8.40

- Best 10% of buildings errors: <1%
- Worst 10% of buildings errors: >19%

Can we identify buildings that will be most/least predictable?

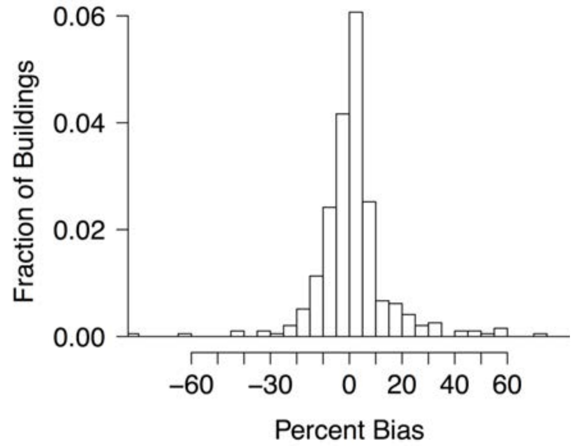
CAN WE SCREEN OR TARGET BUILDINGS TO REDUCE UNCERTAINTY IN M&V?

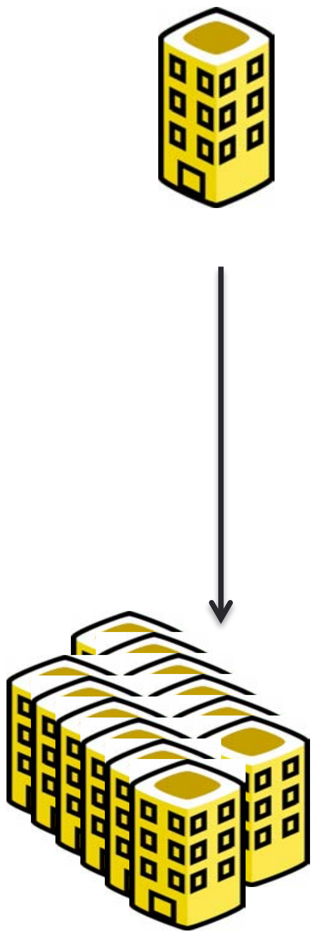
Model	N	10%	25%	50%	75%	90%	Mean
Mean Week	23	3.48	4.10	5.20	5.90	8.32	6.47
Monthly CDD and HDD	72	3.40	4.10	5.45	7.43	9.99	6.82
Day, Time, and Temperature	112	2.70	3.35	4.70	7.55	10.20	6.67
Time of Week and Temperature	110	2.69	3.32	4.55	7.20	10.10	6.33

- No building type was more/less predictable than others (NAICS)
- Simple screening based on training period data reduces errors
- Mean error improves from 8% to 6% , median still ~5%
- In worst 10% of buildings error improves from 19% to ~10%
- In best 10% of buildings error rises (!) from <1% to 2-3%

AGGREGATION OF BUILDINGS REDUCES ERROR TO 1-4%

» Although each savings estimate has error, some are too high and others too low





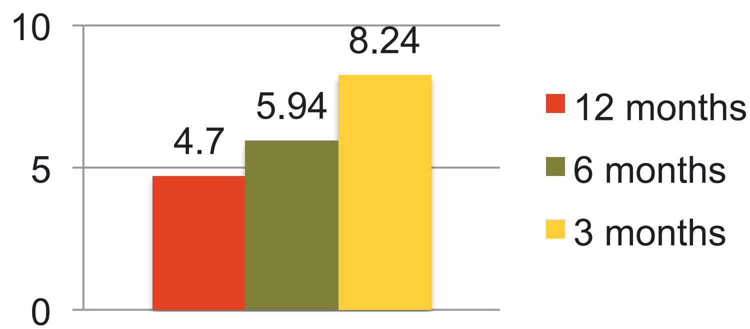
» Aggregation of buildings into a portfolio of ~40 buildings reduces total error to 1-4%

» This reduction in error is not ‘seen’ at the site but *is* at the program level where there is portfolio of participants, reporting at an aggregated level

REDUCING TRAINING FROM 12 TO 6 MONTHS HAS MINIMAL IMPACT ON ACCURACY OF PREDICTIONS

12 months	6 months	3 months
• Current guidance for whole building M&V	• Monthly models fare poorly • No significant degradation in mean, median accuracy • Large increase in error in worst 10% of buildings	• Significant degradation in accuracy • Differences in performance between baseline models appear

median % error for day and change point model



*May be opportunities to shorten M&V for **portfolios**, if willing to tolerate lower site-level accuracy*

KEY TAKEAWAYS - CONCLUSIONS

» LBNL has a way to quantify accuracy of *fully automated* M&V, and identified key metrics

» We have established *performance benchmarks* based on industry standard models

- These benchmarks can be used to set performance *criteria* based on programmatic needs

* Test dataset must be applicable to use context

KEY TAKEAWAYS - CONCLUSIONS

» With interval data, > 12-month training may be possible for whole-building savings estimation

» Median model errors <5%, for 25th percentile <2%, across hundreds of buildings

- no such accuracy prediction is available for engineering calculations

» Depending on required confidence, depth of expected savings, M&V may be able to be conducted in a fully automated manner, or with some engineering intervention

» Promise to scale M&V, unlock deeper savings through multi-measure programs quantifiable at whole-building level

UTILITY INTEREST

» PG&E-ET funded Whole-Building Savings Estimation project by LBNL & QuEST:

- Developed procedure to test accuracy of emerging tools, baseline models for whole-building M&V
- Developed specific testing protocols with ‘blinds’ to protect customer data and vendor IP
- Protocols and test methods used to prequalify tools for inclusion in 2013-2014 Whole Building pilot, 20% multi-measure savings target

PGE Team: Leo Carillo, Mananya Chansanchai, Mangesh Basarkar, Ken Gillespie

» CEE whole buildings committee, key metrics and acceptance criteria for prequalification of models/tools for streamlined delivery of whole-building focused programs

Looking Forward

WHAT ARE WE DOING GOING FORWARD?

- Engage broad group of stakeholders at national level to
  - Gauge conceptual buy-in, need for standard, objective test methods
  - Elicit feedback and vetting of technical aspects of work (TAG participation)
- Extend methodology beyond whole building savings
  - Isolated measures (IPMVP Option B)
  - DR savings
- Use methodology to demonstrate accuracy, compare and contrast new unique models/tools M&V (July solicitation)
- Publish results and models for use, demonstrate with utilities and owners for increased adoption in efficiency community

RFP: ASSESSING ACCURACY OF EMERGING M&V METHODS

- » Request was for unique baseline energy use prediction models from developers
  - LBNL will apply existing statistical methodology to assess performance measurements of savings for building energy efficiency projects and demonstrate model accuracy
  - More info: https://sites.google.com/a/lbl.gov/advancedmandv/

- » Overview of model types selected for evaluation:
  - Gaussian Process Model (GPM)
  - Gaussian Mixture Model (GMM)
  - Neutral Network
  - Regression
    - Advanced Regression with drift
    - Advanced Regression and Nearest Neighbor
  - Combinations:
    - Regression – Bin – Ensemble
    - Bin - Principle competent analysis
    - Ensemble - Multivariate Adaptive Regression Splines

METRICS OF FOCUS

Total normalized bias

$$Total\,Normalized\,Bias = \sum_{i}^{N} \frac{\left(y_i - \hat{y}_i\right)}{y_i} \times 100$$

- Percent difference between total model-predicted energy use and total actual energy use
- Clear relevance to errors in reported savings
- Normalization aids in simultaneous treatment of both large and small building loads
- Bias retains directionality of differences, i.e., under or over-prediction, which has implications for savings payouts and incentives

Coefficient of variation of the root mean squared error

$$CV\,Root\,Mean\,Squared\,Error = \frac{\sqrt{\frac{1}{N}\sum_{i}^{N}\left(y_i - \hat{y}_i\right)^2}}{\bar{y}} \times 100$$

- Squares difference between predictions and data to highlight large differences between predictions and data
- Favors models that predict the overall shape of the energy meter time series
  - Added insight for extrapolation as in normalized savings calculations
- Prominent in industry references such as Guideline 14

CALL FOR DATA!

» Seeking real-world energy use and independent variable data to contribute to
  - removing barriers through evaluation of method accuracy and reliability
  - advancement and scaled adoption of M&V through Automated M&V, or M&V 2.0

» Ideally data includes interval meter data, zip code, and NAICS code
  - 24 months of data (preferred history)
  - Hourly or sub-hourly time intervals
  - Not currently part of 'known' efficiency project

Data use is for research purposes only -- will not be published or shared with third parties

Thank You!

Questions?

Jessica Granderson
JGranderson@lbl.gov
510-486-6792

Itron

ItronUtilityWeek | 2014
resourcefulness in action

WE WANT YOUR FEEDBACK

THERE ARE 2 WAYS TO PROVIDE FEEDBACK ON THIS SESSION

MOBILE APP OR EVALUATION FORMS